1-2021

Reproducibility and Discriminant Validity of the Posterior Shoulder Endurance Test in Healthy and Painful Populations

Neil A. Evans
Ohio University

Suzanne Konz
Marshall University

Arthur J. Nitz
University of Kentucky, arthur.nitz@uky.edu

Timothy L. Uhl
University of Kentucky, tluhl2@uky.edu

Reproducibility and Discriminant Validity of the Posterior Shoulder Endurance Test in Healthy and Painful Populations

Digital Object Identifier (DOI)
https://doi.org/10.1016/j.ptsp.2020.10.014

Notes/Citation Information
Published in *Physical Therapy in Sport*, v. 47.

Reproducibility and Discriminant Validity of the Posterior Shoulder Endurance Test in Healthy and Painful Populations

Acknowledgments: The authors would like to acknowledge the physical therapy students that assisted as research assistants during data collection as part of their capstone requirement for their Doctor of Physical Therapy degrees. We also acknowledge the [redacted] for helping to provide the facilities and resources to conduct this research.

ABSTRACT

Objective: This investigation measured the reproducibility and discriminant validity of the Posterior Shoulder Endurance Test (PSET) on painful and non-painful populations.

Design: Reliability and validity study

Setting: Laboratory setting

Participant: Thirty subjects (male=11; female=19)

Main Outcome Measures: Time to failure (TTF) was the primary outcome measure to determine reliability of the PSET. Discriminant validity identified with receiver operator characteristic (ROC) curves utilized TTF separately in men and women since they used different loads.

Results: There were 25/30 subjects (painful=12; non-painful=13) tested a second time. ICC, SEM, and MDC_{90} ranged respectively from 0.77, 13.1 seconds, 30.6 seconds in the painful group to 0.85, 7.3 seconds, 17 seconds in the non-painful group. The male ROC curve AUC was 0.833 with 47 seconds resulting in the best combination of sensitivity = 0.833, and specificity = 0.80. The female ROC curve AUC was 0.633 with 46 seconds resulting in the best combination of sensitivity = 0.600 and specificity = 0.889 at 46 seconds.

Conclusion: The PSET is a reliable way to measure shoulder girdle muscular endurance. These data suggest that the PSET discriminates painful and non-painful individuals better in men compared to women.

Key Words: Shoulder endurance, test re-test reliability, discriminant validity

1. INTRODUCTION

Non-traumatic shoulder pain accounts for 44-65% of all musculoskeletal shoulder complaints (Lopes, Timmons, Grover, Ciconelli, & Michener, 2015; van der Windt, Koes, de Jong, & Bouter, 1995; Vecchio, Kavanagh, Hazleman, & King, 1995). The mechanisms leading to non-traumatic shoulder pain are multifactorial, including kinematic alterations (Ludewig & Reynolds, 2009; Phadke & Ludewig, 2013), anatomic variations (Ludewig & Reynolds, 2009), intrinsic tendon degeneration (Michener, McClure, & Karduna, 2003), and a lack of muscular endurance (Chopp-Hurley, O'Neill, McDonald, Maciukiewicz, & Dickerson, 2015; Chopp, O'Neill, Hurley, & Dickerson, 2010; Lopes, et al., 2015; Michener, et al., 2003; Seitz, McClure, Finucane, Boardman, & Michener, 2011). Therefore, clinicians are challenged to differentiate between various mechanistic contributors. Since the prevalence of non-traumatic shoulder pain is high, fully understanding the role of each potential contributing factor becomes very important for clinicians to individualize treatment.

Muscular endurance is the ability of the muscle to produce force over an extended time or perform multiple repetitions of a load (Backman, Johansson, Hager, Sjoblom, & Henriksson, 1995). Whereas, muscular strength is the ability to produce a maximal amount of force for a short period. Therefore, the assessment methods for muscular endurance should be unique when compared to assessments of strength. Additionally, sports-related overuse shoulder injuries such as rotator cuff tendinopathy and alike caused by overhead activity, have

been attributed to a lack of muscle endurance (Chopp-Hurley, et al., 2015; Michener, et al., 2003; Sein, et al., 2010; Seitz, et al., 2011). Madsen, et al. (Madsen, Badault, & Nybo, 2018) found that youth badminton players with greater muscular endurance had improvements in performance. Yet, clinical tests focusing on shoulder muscular endurance are scarce (Day, Bush, Nitz, & Uhl, 2015; Edmondston, et al., 2008; Kumta, MacDermid, Mehta, & Stratford, 2012). Available muscular endurance measures either lack shoulder girdle specificity (Kumta, et al., 2012), or lack clinical measurement properties necessary prior to implementation (Day, et al., 2015; Edmondston, et al., 2008). Reliability and validity of clinical measures are important to assure the efficacy of the clinical tool (Impellizzeri & Marcora, 2009).

The Scapular Endurance Test (SET) and the Posterior Shoulder Endurance Test (PSET) are two clinical tests found in the literature, which measure shoulder girdle muscular endurance (Edmondston, et al., 2008; Evans, Dressler, & Uhl, 2018). The SET is described as having a client face a wall with the shoulders and elbows flexed to 90° with a spacer positioned between the elbows (Edmondston, et al., 2008). The client holds the spacer with the elbows while maintaining a 1 Kg load cell between his/her hands by performing shoulder external rotation (Edmondston, et al., 2008). The serratus anterior muscle is thought to be a primary muscle in the SET due to the test position, but muscle activity was not recorded (Edmondston, et al., 2008). Without knowing the extent of scapular muscle action, the SET is lacking as a clinical measure of scapular muscular

endurance. The PSET is an isometric test performed to failure (Evans, et al., 2018). Individuals hold a standardized external load based on body weight and arm length while lying prone with the shoulder in 90° of horizontal abduction and full external rotation (Chaffin DB, 1999; Evans, et al., 2018). Time to task failure (TTF) of 58.1 seconds (95%CI = 57.3-58.9) in asymptomatic females and 68.5 seconds (95%CI = 67.4- 69.6) in asymptomatic males has been reported (Evans, et al., 2018). However, the TTF in painful populations has not been investigated and would provide clinicians with another test to assess muscle performance. Unlike the SET, electromyography results have been reported for the PSET. The PSET fatigues the upper, middle, and lower fibers of the trapezius, the infraspinatus, and the posterior deltoid at a similar rate between all muscles tested with one exception in males and females (Evans, et al., 2018). Day et al. (Day, et al., 2015) demonstrated muscular endurance deficits in patients with lateral epicondylalgia and those without in a variation of the PSET, suggesting the test can identify muscular performance deficits that should be addressed during rehabilitation. Based on previous findings, the PSET shows promise as a measure of shoulder muscular endurance for clinicians to address during rehabilitation. However, reliability and discriminant validity has yet to be evaluated in painful and non-painful populations for the PSET.

Therefore, the purpose of this study was to determine the inter-day test re-test reliability of the PSET in both non-painful individuals and individuals with stable shoulder pain. A second purpose was to evaluate the discriminant validity

of the PSET in males and females separately with and without shoulder pain. Since males and females held differing amounts of external load, the discriminant validity of the TTF determined from the PSET could not be directly compared between sexes. Discriminant validity, as measured by a receiver operator characteristic (ROC) curve, will assist in establishing the diagnostic accuracy of the PSET and developing a cut-off score. We hypothesize the PSET will have moderate to excellent reliability (ICC > 0.70) in painful and non-painful individuals and be able to differentiate painful individuals from non-painful individuals.

2. METHODS

2.1. Evaluators

There were two evaluators in this study. The primary investigator is a physical therapist with 16 years of experience. The primary investigator was not involved in screening the subjects to determine if they met the criteria for being in the painful group. Additionally, group classification remained blinded until after all data were collected. The primary investigator performed all PSET testing. The secondary investigator was a second-year physical therapy student that had been trained in class as well as by the primary investigator and had successfully passed skill checks for the special tests. The secondary investigator was responsible for obtaining informed consent, screening the subject for inclusion into the painful group, and any follow-up with the subject after testing. This investigator was also responsible for matching painful subjects with non-painful subjects based on age

and maintaining left/right the side being tested between groups. Since painful subjects always had their involved shoulder tested, the side tested on the non-painful subjects were also controlled to match the number of painful subjects tested. The primary investigator trained the secondary investigator on all clinical testing used for determining subject inclusion prior to initiating data collection.

2.2. Subjects

Subjects between 20-60 years of age, with and without non-traumatic shoulder pain, were recruited to participate in the study. All subjects completed university-approved informed consent, demographic information, and the Pennsylvania Shoulder Score (PSS) (Leggin, et al., 2006). The secondary investigator recorded weight, height, arm length, and in order to calculate standard external torque applied to a subject's arm (Chaffin DB, 1999; Evans, et al., 2018). The secondary investigator also performed clinical testing including Hawkins-Kennedy, Empty can, Neer' Impingement sign, a painful arc between 60-120°, and pain with external rotation manual muscle testing as described by Michener et al. (Michener, Walsworth, Doukas, & Murphy, 2009). Subjects were excluded if they had uncontrolled hypertension, glaucoma, or a neurological diagnosis that would impede them from performing the test.

Non-painful subjects were recruited by word-of-mouth from local orthopedic physician offices and rehabilitation clinics when they were being seen for non-shoulder conditions. Non-painful subjects had no history of shoulder surgery nor a history of any other type of surgery within the last 6 months prior to testing.

Additional inclusion for the non-painful subjects was a score >90/100 on the PSS and had ≤ 2 positive clinical tests performed by the secondary investigator.

Painful subjects were recruited by word-of-mouth from local orthopedic physician offices and physical therapy clinics and the community-at-large. Inclusion criteria for the painful group included a PSS score of < 90/100 and three of the following tests being positive: Hawkins-Kennedy, Empty can, Neer' Impingement sign, a painful arc between 60-120°, and pain with external rotation manual muscle testing (Michener, et al., 2009). Three of the five positive tests indicated a significant area under the curve (AUC) = 0.79 ($p<0.01$) with a positive likelihood ratio (LR+) post-test probability of 54.40% (Michener, et al., 2009), demonstrating a high likelihood that individuals have the condition. Subjects in the painful group were excluded if found to have a positive finding for abduction drop arm test, external rotation lag sign, or a positive lift-off test due to the high likelihood of a rotator cuff tear (Cook & Hegedus, 2013).

2.3. Procedure

The arm length, measured from the lateral border of the acromion to the distal end of the radial styloid process, and body weight were used for standardizing the external toque across participants (Evans, et al., 2018). The external torque was standardized based on published anthropometric data using the 50th percentile for males and females (Chaffin DB, 1999). After using anthropometric data to estimate torque provided by the arm alone, an additional external load was provided to the nearest 0.23 kg resulting in external torque for

155 males equaling 21 ± 2 Nm and the external torque for females equaling 13 ± 1
156 Nm. The range of weight held for the male subjects was between 1.36 – 2.5 kg.
157 Female subject external loads ranged between 1.14 – and 1.59 kg. These ranges
158 were similar to previously published values ranging from 2.05 -2.5kg in males and
159 1.36 – 1.59 kg in females (Evans, et al., 2018).

160 Subjects performed a five-minute warm-up on an upper-body ergometer to
161 minimize the risk for muscular strain. After subjects were familiarized with the
162 testing procedures, s/he laid prone and held the arm at 90° of horizontal abduction
163 against a stand-alone target to ensure proper form was maintained (**FIGURE 1**).

164

FIGURE 1. PSET testing position. The shoulder is at 90˚ of abduction and 90˚ of horizontal abduction.

The TTF was measured with a stopwatch and recorded as the time (seconds) in which the participant initially contacted the target until the participant could no longer maintain contact with the target. Verbal encouragement was provided throughout the testing procedure. Failure of testing was defined as not maintaining form or consistent contact with the target. Examples of test failure behavior included excessive trunk rotation, inability to maintain contact with the stand-alone target after verbal encouragement was provided, or self-selected stoppage. Painful subjects were educated prior to testing to stop if the pain became intolerable. The secondary investigator interviewed each subject after testing to determine why the activity was stopped. Three painful subjects discontinued the testing secondary to an increase in pain. However, the TTF for those three subjects was obtained as previously described. The exact procedure was reproduced 7-10 days later to assess test re-test reliability. To ensure no change in symptoms between testing days, the secondary investigator asked the subjects to rate any change in symptoms using the global rate of change score (GROC) (Stevens, et al., 2019). The GROC is an 11-point scale measuring a patient's perceived improvement or deterioration (Stevens, et al., 2019). Subjects were permitted to participate in the second day of testing if they reported scores of -1, 0, or +1. Test-re-test reliability of the GROC ranges between ICC = 0.90-0.99 (Stevens, et al., 2019). Three out of fifteen subjects reported a negative change in the GROC score that exceeded -1 and were excluded from the reliability portion of

the study. It is unknown whether the negative change in the GROC score was due to the testing or was simply due to an exacerbation in their symptoms.

2.4. Statistical Analysis

Prior to determining the reliability, a Shapiro-Wilk test ensured the TTF across groups was normally distributed ($p > 0.05$). TTF was used to assess the inter-day reliability of the PSET. Interclass correlation coefficients ($ICC_{2,1}$) were calculated for the painful group, non-painful group, and total participants separately. ICCs were considered poor if values were <0.5, moderate if between 0.5 and 0.75, good if between 0.75 and 0.90, and excellent if > 0.90(Koo & Li, 2016). ICCs were used to determine the standard error measurement ($SEM = SD_{pooled} * \sqrt{(1 - ICC)}$), and minimal detectable change at 90% ($MDC_{90} = (SEM * \sqrt{2}) * 1.65$) for total, painful, and non-painful groups.

Separate male and female ROC curves were calculated based on day one testing. The ROC curve coordinates are utilized to determine diagnostic validity, which provides the sensitivity and specificity of a test. The sensitivity of a test is the test's ability to identify a true positive, and specificity is the ability of the test to identify a true negative outcome. The ROC coordinates yielding the best combination of sensitivity and specificity were used to identify a cut-off score for the TTF during the PSET for males and females separately. Cut-off scores should be considered the point at which the test best discriminates individuals likely to have shoulder pain and those likely not to have shoulder pain, therefore, aiding

clinicians in the interpretation of the PSET results (Carter, Pan, Rai, & Galandiuk, 2016; Riddle & Stratford, 1999). The area under the curve (AUC) of the ROC curve provides the likelihood of correctly identifying the condition of true positives and true negatives. Therefore, the diagnostic accuracy of the clinical test can be interpreted as follows: an AUC between 0.90-1.0 = excellent, 0.80-0.90 = good, 0.70-0.80 = moderate, 0.60-0.70 = poor, and < 0.60 = useless (Carter, et al., 2016; Portney & Watkins, 2009).

3. RESULTS

Thirty subjects participated in this study (female=19; male=11). Demographics and PSS are presented in **TABLE 1**. As expected, painful subjects had significantly lower PSS scores than non-painful subjects ($p<0.001$).

Sex	Variable	Treatment Group	Mean ± SD	p-value
Combined	Weight	Non-painful	147.6 ± 23.9	0.069
		Painful	178 ± 55.7	
	Height	Non-painful	168.3 ± 6.7	0.185
		Painful	173 ± 11.2	
	Age	Non-painful	32.9 ± 12.7	0.769
		Painful	34.3 ± 13.0	
	PSS total score	Non-painful	97.9 ± 4.0	<0.001*
		Painful	72.2 ± 13.9	
Males	Weight	Non-painful	168 ± 20.7	0.202
		Painful	210 ± 66.2	
	Height	Non-painful	174.8 ± 6.1	0.008*
		Painful	184.2 ± 2.7	
	Age	Non-painful	33.4 ± 9.3	0.250
		Painful	41.7 ± 12.4	
	PSS total score	Non-painful	98.4 ± 1.5	0.013*
		Painful	70.7 ± 19.8	
Females	Weight	Non-painful	136.3 ± 17.6	0.146
		Painful	158.4 ± 40.0	
	Height	Non-painful	164.7 ± 3.7	0.616
		Painful	166.3 ± 8.5	
	Age	Non-painful	32.6 ± 14.7	0.656
		Painful	29.8 ± 11.7	
	PSS total score	Non-painful	97.7 ± 4.9	<0.001*
		Painful	73.1 ± 10.0	

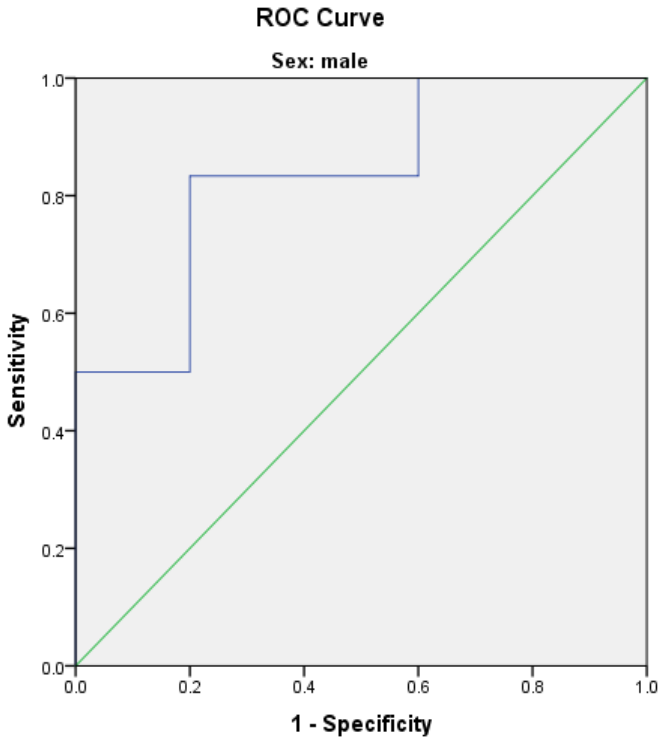
TABLE 1. Patient Demographics. NOTE: Independent t-test compared across groups. * Indicates Significance <0.05. PSS total = Pennsylvania Shoulder Score total score.

Intraclass correlation coefficients ($ICC_{2,1}$) were assessed on 25/30 participants (painful = 12, non-painful =13). One subject in each group had personal conflicts, and three subjects in the painful group had a negative change in GROC score that exceeded the inclusion (**TABLE 2**).

	Total Group	Painful Group	Non-painful Group
TTF Day 1 (Mean ± SD)	51.8 ± 25.5 (n=30)	43.3 ± 27.8 (n=16)	61.5 ± 19.3 (n=14)
TTF Day 2 (Mean ± SD)	55.5 ± 20.8 (n=25)	53.3 ± 24.4 (n=12)	57.6 ± 17.7 (n=13)
$ICC_{2,1}$ (95%CI)	0.80 (0.58 – 0.91)	0.77 (0.40 - 0.93)	0.85 (0.58 – 0.95)
SEM (sec)	10.4	13.1	7.3
MDC90 (sec)	24.4	30.6	17.0

TABLE 2. Intraclass correlation coefficients ($ICC_{2,1}$), standard error measurements (SEM), and minimal detectable changes (MDC_{90}) of the Posterior Shoulder Endurance Test in total, painful, and non-painful populations. NOTE: TTF = Time to Task Failure of the PSET.

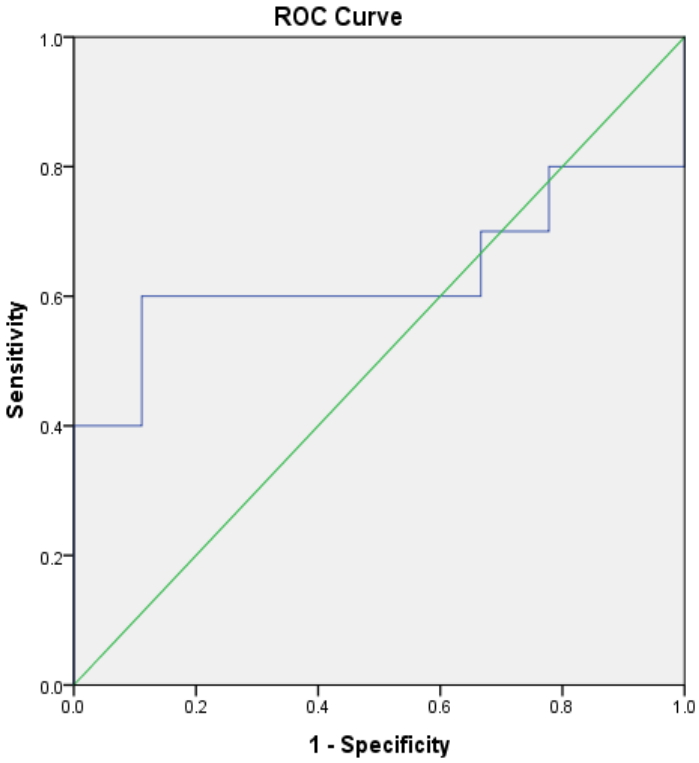
Separate ROC curves were used for male (painful = 6; non-painful = 5) and female (painful = 10; non-painful = 9) participants since different standardized loads were used. Male ROC AUC was 0.833 (CI95%= 0.58-1.0) (**FIGURE 2**). The female ROC AUC was 0.633 (CI95%= 0.361-0.906) (**FIGURE 3**). The male ROC had a sensitivity = 0.833, and specificity = 0.80 at 47 seconds. While the female ROC curve had a sensitivity = 0.600 and specificity = 0.889 at 46 seconds.



244

245 **FIGURE 2**. Male Participant Receiver Operating Characteristic Curve (ROC) of the
246 PSET time to task failure.

247



248

249 **FIGURE 3**. Female Participant Receiver Operating Characteristic Curve (ROC) of
250 the PSET time to task failure.

251

4. DISCUSSION

The primary purpose of this investigation was to determine the clinical utility of the PSET by examining the inter-day reliability and discriminant validity of the measure. The data suggest the PSET has good reliability in non-painful populations ($ICC_{2,1}$ = 0.85), and in painful populations ($ICC_{2,1}$ = 0.77). Although the subjects denied changes in symptoms from day one to day two using the GROC, sub-clinical symptom changes may contribute to the reduction in reliability observed in the painful group. Since reliability is measuring the stability of the test, any symptoms must be consistent, or the test performance might change (Portney & Watkins, 2009). Therefore, individuals with pain would be more susceptible to labile symptoms, thus producing lower ICC values. However, since ICC values of >0.75 have been reported as good reliability scores (Koo & Li, 2016; Portney & Watkins, 2009), and in the absence of other clinical measures for posterior shoulder girdle muscular endurance in subjects with and without shoulder pain, these ICCs should be considered an acceptable level of reliability.

The minimal detectable change (MDC) is a distribution-based value influenced by the measurement error of a test, which is directly influenced by a test ICC or stability of a test. Therefore, as the reliability decreases, the responsiveness of the measure would decrease, and the ability of a test to demonstrate a real change requires a greater change in the measured value. Based on the current data, to be 90% confident that a true change in TTF of the PSET occurred, there should be 17 seconds change in a non-painful population

and 31 seconds change in a painful population. In a similar testing procedure, the MDC_{90} of the PSET at 135˚ isometric shoulder abduction was 24 seconds (Day, 2013). While Day's findings were slightly higher than this current study, the possibility of a learning effect would have likely inflated the MDC value. So, the MDC value of 17 seconds in this current study for a non-painful population is reasonable. The PSET MDC in a painful shoulder population has not been reported in previous literature. Based on the current and a previous study (Day, 2013), clinicians should consider using 30 seconds to represent a functional improvement in a painful population and 17 seconds in a non-painful population when using the PSET as a measurement tool for posterior shoulder endurance.

The scapular endurance test (SET) described by Edmondston et al. (Edmondston, et al., 2008) reported reliability of 0.67 (CI95% = 0.31-0.85) with an MDC_{95} of 30.1 seconds in individuals with neck pain. The reliability of the scapular endurance test in a healthy population has not been reported. Additionally, the SET was only tested on individuals with neck pain, not shoulder pain. While the SET is performed until failure, the muscles responsible for the activity likely differ from the PSET. The muscles fatiguing during the SET have not been investigated (Edmondston, et al., 2008). However, Elkstrom et al. (Ekstrom, Donatelli, & Soderberg, 2003) described a similar movement as demonstrating high muscle activity of serratus anterior and trapezius. Conversely, Evans et al. (Evans, et al., 2018) found that muscle activity is fatiguing in the trapezius, infraspinatus, and posterior deltoid during the PSET. So, while the results of this

297 investigation demonstrate comparable reliability and MDC values to the SET in
298 painful populations, the PSET offers unique information since the scapular position
299 and muscles being fatigued differ.

300 The second purpose of this investigation was to determine if the TTF was
301 able to differentiate individuals with and without shoulder pain. Discriminant
302 validity is particularly important in a clinical setting, as clinicians are evaluating
303 patient symptoms. ROC curve plots help determine the clinical utility by plotting
304 true positive findings (Sensitivity) against false positives (1-Specificity)(Carter, et
305 al., 2016). The current results support the PSET is good for discriminating males
306 with and without shoulder pain (AUC=0.883), but poor at discriminating females
307 with and without shoulder pain (AUC=0.633)(Carter, et al., 2016). Upon closer
308 examination of the data, there was one painful female subject that held the PSET
309 for 102 seconds, thus skewing the sensitivity and specificity of the female graph. If
310 the ROC curve were performed without the one outlier, the AUC= 0.704 (CI 95%
311 0.439, 0.969) would have improved to a moderate level. Therefore, one subject
312 made a significant difference in the ROC curve due to the small sample size.
313 Since the sample size was limited for both sexes, the authors feel further research
314 is warranted to confirm or refute these results.

315 The ROC curve can also establish the point at which the TTF has the best
316 combination of true positives and true negatives, known as a cut-off score (Carter,
317 et al., 2016). These data demonstrate a cut-off score that differentiated those with
318 and without shoulder pain of 46 and 47 seconds in females and males,

respectively. The cut-off score can be interpreted as the time used to differentiate those with shoulder pain from those without shoulder pain. In a perfect test, individuals without pain should score higher than the cut-off time, and individuals with pain should score below the cut-off time. The cut-off score of 46 seconds resulted in correctly classifying 75% (8/12) of the non-painful and 86% (6/7) of the painful female participants. Therefore, the specificity is much higher compared to the sensitivity in the female population. Similarly, the male ROC curve identified a cut-off score of 47 seconds resulting in correctly classifying 80% (4/5) of the non-painful and 83% (5/6) of the painful male participants. The combination of sensitivity (0.833) and specificity (0.80) in the male cohort produced a more meaningful combination. While these findings are novel, further research needs to be performed to improve the precision of the cut-off scores.

The PSS was collected from all participants and used to discriminate between the painful and non-painful participants (**TABLE 1**). However, the average PSS for the painful group was still relatively high in this sample, which indicates that they had a relatively high function and satisfaction with low pain levels. Therefore, individuals with more significant amounts of pain or acute injury may not be able to tolerate the PSET testing position. Since pain can limit performance on any functional test, a clinician should consider adding the PSET after pain severity has been mitigated. The current painful sample had an average PSS pain subscale of 20.6 ± 3.9 out of a score of 30, where a score of 30/30 would represent no pain. Using this sample as a guide, clinicians should be able

to reasonably test patients with the PSET if they score ≥17/30 on the PSS pain subscale or a similar construct (Leggin, et al., 2006). More research is needed to determine if an increase in pain and functional loss limits the subject's ability to perform the PSET.

4.1. Limitations

Despite all attempts to limit the extraneous factors influencing our results, this study is not without limitations. A limitation of the current investigation is the sample size. A larger number of participants reduces the likelihood of over-estimation or under-estimation of both reliability and validity measures (Portney & Watkins, 2009). The results of this investigation should be used cautiously until further evidence either supports or refutes its findings.

Since the results of this study are dependent on maximal effort performance by subjects, the authors cannot assure that all participants were performing maximally. There was an underlying assumption that all subjects would give maximal effort, and clear instructions and expectations of testing were provided to participants prior to testing. However, multiple factors might cause an individual to stop the test including muscular fatigue, pain, or lack of motivation. A clear definition of test failure was implemented to mitigate participants ceasing the PSET without maximal effort. Yet, three subjects reported stopping the testing secondary to pain, with the remaining participants demonstrated test failure as defined *a priori*. A second limitation regarding effort dependent testing is whether

the fatigue is of central or peripheral origin (Enoka & Duchateau, 2008). In studies using human subjects, it is difficult to control for the type of fatigue occurring.

Lastly, the generalizability of this study to a painful population may be limited. Inclusion criteria were set to assure a strong likelihood that the painful group had chronic pain that resembled tendinopathy without evidence of a tendon tear (Michener, et al., 2009). Although individuals with and without shoulder pain were included in this study, only sixty-nine percent of the painful subjects in the current study were seeking medical care. Therefore, the results of this study may not represent a population that typically seeks medical care. At this time, the authors suggest implementing the PSET after acute pain and dysfunction have subsided.

5. CONCLUSION

The PSET is a muscular endurance clinical measure targeting the posterior shoulder girdle. The study supports the PSET is a reliable tool for measuring posterior shoulder muscle endurance in painful and non-painful populations (ICC = 0.77- 0.85). The PSET discriminant validity was stronger in the male population than the female population. Clinicians can use cut-off scores of 46 and 47 seconds in females and males, respectively, to help determine if muscular endurance is contributing to shoulder pain. The PSET's minimal detectable change score of 17 and 31 seconds for non-painful and painful populations, respectively, help clinicians measure change after an intervention. More research

should be performed to overcome the limitations of the current study and establish a more robust diagnostic validity of the PSET. Future research should determine the minimally clinically important difference (MCID) of the PSET to improve responsiveness measures and if an increase in TTF of the PSET equates to an improvement in painful symptoms.

REFERENCES

Backman, E., Johansson, V., Hager, B., Sjoblom, P., & Henriksson, K. G. (1995). Isometric muscle strength and muscular endurance in normal persons aged between 17 and 70 years. *Scandinavian Journal of Rehabilitation Medicine, 27*, 109-117.

Carter, J. V., Pan, J., Rai, S. N., & Galandiuk, S. (2016). ROC-ing along: Evaluation and interpretation of receiver operating characteristic curves. *Surgery, 159*, 1638-1645.

Chaffin DB, A. G., Martin BJ. (1999). *Occupational Biomechanics*. New York, NY: John Wiley & Sons, Inc.

Chopp-Hurley, J. N., O'Neill, J. M., McDonald, A. C., Maciukiewicz, J. M., & Dickerson, C. R. (2015). Fatigue-induced glenohumeral and scapulothoracic kinematic variability: Implications for subacromial space reduction. *Journal of Electromyography and Kinesiology*.

Chopp, J. N., O'Neill, J. M., Hurley, K., & Dickerson, C. R. (2010). Superior humeral head migration occurs after a protocol designed to fatigue the rotator cuff: a radiographic analysis. *Journal of Shoulder and Elbow Surgery, 19*, 1137-1144.

Cook, C., & Hegedus, E. J. (2013). *Orthopedic Physical Examination Tests: An Evidence-Based Approach* (2nd ed.). Upper Saddle River, NJ: Pearson Education Inc.

Day, J. M. (2013). *Scapular Muscle Assessment in Patients with Lateral Epicondylagia.* University of Kentucky, UKnowledge.

Day, J. M., Bush, H., Nitz, A. J., & Uhl, T. L. (2015). Scapular muscle performance in individuals with lateral epicondylalgia. *Journal of Orthopaedic and Sports Physical Therapy, 45*, 414-424.

Edmondston, S. J., Wallumrod, M. E., Macleid, F., Kvamme, L. S., Joebges, S., & Brabham, G. C. (2008). Reliability of isometric muscle endurance tests in subjects with postural neck pain. *Journal of Manipulative and Physiological Therapeutics, 31*, 348-354.

Ekstrom, R. A., Donatelli, R. A., & Soderberg, G. L. (2003). Surface electromyographic analysis of exercises for the trapezius and serratus anterior muscles. *Journal of Orthopaedic and Sports Physical Therapy, 33*, 247-258.

Enoka, R. M., & Duchateau, J. (2008). Muscle fatigue: what, why and how it influences muscle function. *Journal of Physiology, 586*, 11-23.

Evans, N. A., Dressler, E., & Uhl, T. (2018). An electromyography study of muscular endurance during the posterior shoulder endurance test. *Journal of Electromyography and Kinesiology, 41*, 132-138.

Impellizzeri, F. M., & Marcora, S. M. (2009). Test Validation in Sport Physiology: Lessons Learned From Clinimetrics. *International Journal of Sports Physiology and Performance, 4*, 269-277.

Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine, 15*, 155-163.

Kumta, P., MacDermid, J. C., Mehta, S. P., & Stratford, P. W. (2012). The FIT-HaNSA demonstrates reliability and convergent validity of functional performance in patients with shoulder disorders. *Journal of Orthopaedic and Sports Physical Therapy, 42*, 455-464.

Leggin, B. G., Michener, L. A., Shaffer, M. A., Brenneman, S. K., Iannotti, J. P., & Williams, G. R., Jr. (2006). The Penn shoulder score: reliability and validity. *Journal of Orthopaedic and Sports Physical Therapy, 36*, 138-151.

Lopes, A. D., Timmons, M. K., Grover, M., Ciconelli, R. M., & Michener, L. A. (2015). Visual scapular dyskinesis: kinematics and muscle activity alterations in patients with subacromial impingement syndrome. *Archives of Physical Medicine and Rehabilitation, 96*, 298-306.

Ludewig, P. M., & Reynolds, J. F. (2009). The association of scapular kinematics and glenohumeral joint pathologies. *Journal of Orthopaedic and Sports Physical Therapy, 39*, 90-104.

Madsen, C. M., Badault, B., & Nybo, L. (2018). Cross-Sectional and Longitudinal Examination of Exercise Capacity in Elite Youth Badminton Players. *Journal of Strength and Conditioning Research, 32*, 1754-1761.

Michener, L. A., McClure, P. W., & Karduna, A. R. (2003). Anatomical and biomechanical mechanisms of subacromial impingement syndrome. *Clinical Biomechanics (Bristol, Avon), 18*, 369-379.

Michener, L. A., Walsworth, M. K., Doukas, W. C., & Murphy, K. P. (2009). Reliability and diagnostic accuracy of 5 physical examination tests and combination of tests for subacromial impingement. *Archives of Physical Medicine and Rehabilitation, 90*, 1898-1903.

Phadke, V., & Ludewig, P. M. (2013). Study of the scapular muscle latency and deactivation time in people with and without shoulder impingement. *Journal of Electromyography and Kinesiology, 23*, 469-475.

Portney, L., & Watkins, M. (2009). *Foundations of Clinical Research: Application to Practice* (3rd ed.). Upper Saddle River, New Jersey: Pearson Prentice Hall.

Riddle, D. L., & Stratford, P. W. (1999). Interpreting Validity Indexes for Diagnostic Tests: An Illustration Using the Berg Balance Test. *Physical Therapy, 79*, 939.

Sein, M. L., Walton, J., Linklater, J., Appleyard, R., Kirkbride, B., Kuah, D., & Murrell, G. A. (2010). Shoulder pain in elite swimmers: primarily due to swim-volume-induced supraspinatus tendinopathy. *British Journal of Sports Medicine, 44*, 105-113.

Seitz, A. L., McClure, P. W., Finucane, S., Boardman, N. D., 3rd, & Michener, L. A. (2011). Mechanisms of rotator cuff tendinopathy: intrinsic, extrinsic, or both? *Clinical Biomechanics (Bristol, Avon), 26*, 1-12.

Stevens, M. L., Lin, C. C., van der Ploeg, H. P., De Sousa, M., Castle, J., Nicholas, M. K., & Maher, C. G. (2019). Feasibility, Validity, and Responsiveness of Self-Report and Objective Measures of Physical Activity in Patients With Chronic Pain. *Pm r*.

van der Windt, D. A., Koes, B. W., de Jong, B. A., & Bouter, L. M. (1995). Shoulder disorders in general practice: incidence, patient characteristics, and management. *Annals of the Rheumatic Diseases, 54*, 959-964.

Vecchio, P. C., Kavanagh, R. T., Hazleman, B. L., & King, R. H. (1995). Community survey of shoulder disorders in the elderly to assess the natural history and effects of treatment. *Annals of the Rheumatic Diseases, 54*, 152-154.